Geophys. J. Int. (1992) **108,** 865–882

Spectra of mantle shear wave velocity structure

J. Huw Davies,* Olafur Gudmundsson,† and R. W. Clayton
Seismological Laboratory, 252-21 California Institute of Technology, Pasadena, CA 91125, USA

Accepted 1991 September 26. Received 1991 May 21

SUMMARY

We applied the stochastic method of Gudmundsson, Davies & Clayton (1990) (which was applied to ISC *P*-wave data) to teleseismic ISC *S*-wave data to obtain an independent estimate of mantle structure. We inverted the variance of *S*-wave traveltime residuals of bundles of rays to obtain a description of the spectrum of lateral heterogeneity as a function of depth through the mantle. The technique yields robust estimates of the traveltime scattering power (the product of a characteristic scalelength of heterogeneity and the mean square of slowness perturbations). We can estimate the characteristic scalelength (half-width), from the autocovariance; which can be reconstructed from the spectra. Hence by division, we can estimate the root mean square slowness. By extrapolating the variance of bundles of rays to bundles of zero cross-sectional area we can also estimate the scale-incoherent signal (which is a plausible estimate of the noise in the data), which is removed from the data.

We find that most of the structure generating shear wave traveltime residuals is located in the uppermost mantle. About half of the structure is short scale (harmonic degree $l > 50$). The large-scale structure ($l < 50$) has a half-width of about 500 km in the upper half of the mantle. This *S*-wave half-width is consistent with the *P*-wave half-widths determined by Gudmundsson *et al.* (1990). The *S*-wave half-width in the lower half of the mantle is poorly constrained. It varies from 500 to 3000 km, which spans the better constrained value of 1200 km found by Gudmundsson *et al.* (1990) for *P*-waves. The incoherent scatter suggests that the signal-to-noise ratio of the *S*-wave data set is around 1.5.

Assuming that the compressional and shear wave velocity variations are correlated then the signal weighted value of the ratio $d \ln (V_s)/d \ln (V_p)$ is ≈ 2, as also found in normal mode studies. This is much larger than the value of ≈ 0.8–1.4 suggested by laboratory experiments undertaken at atmospheric pressure. There is no evidence of periodicity in the traveltime autocovariance; this suggests little or no periodicity in the underlying convection. The short half-width through most of the mantle suggests high Rayleigh number convection, with its attendant small-scale structures. The power decreases by an order of magnitude or more in going from the upper mantle to the lower mantle, the same as found by Gudmundsson *et al.* (1990) for *P*-waves. This large difference suggests either a change in convective regime and/or a difference in the temperature sensitivity of elastic constants in both layers. The increased short-scale structure at the top of the mantle suggests that a large part of the seismic signature at this boundary is compositional, since one would expect a red spectrum for a thermal boundary layer. The derived spectra between $l \approx 10$ and $l \approx 50$ are similar in shape to spectra from the mantle convection simulations of Glatzmaier (1988) with a Rayleigh number of 10^6–10^7, which would suggest layered convection, if the comparison is valid.

Key words: body waves, inversion, stochastic, traveltime variance.

* Presently at: Institute of Theoretical Geophysics, Department of Earth Sciences, University of Cambridge, Downing St., Cambridge CB2 3EQ, UK.
† Presently at: Research School of Earth Sciences, Australian National University, GPO Box 4, Canberra, ACT 2601, Australia.

1 INTRODUCTION

Since the inception of plate tectonics, we have been unable to elucidate the details of how it is powered by mantle convection. With the advent of deterministic (tomographic) models of lateral seismic variations (Dziewonski 1984; Clayton & Comer 1983; Hager & Clayton 1989) we have the first chance of imaging directly the driving forces of mantle convection and hence constraining mantle convection and establishing how plate tectonics is powered. The body wave tomography models quoted above have in principle a higher spatial resolution than surface wave or free oscillation studies because of the shorter wavelengths of body waves. However, current lower mantle body wave studies suffer from poor variance reduction, and there are many questions related to the reliability of these tomographic models, including the following. How much random noise is there in the data and how well is it eliminated from the model? How much aliasing is there, due to the finite parametrization [Dziewonski (1984), $l \leq 6$; Clayton & Comer (1983), cell size ≥ 250 km, $l \leq 36$]? How well do the station corrections account for the crustal and upper mantle structure? How are the results affected by the uneven sampling?

Gudmundsson *et al.* (1990), hereafter referred to as GDC, presented a theory to invert the scatter in the traveltimes of seismic waves to obtain the spectra of lateral seismic velocity variations through the mantle. Following Davies, Gudmundsson & Clayton (1988), we also estimate the error in the ISC data set. This estimated error is subtracted from the data and does not enter the model, while the model parametrization is extended to the smallest scale avoiding aliasing. The method models the whole mantle and attempts to account for the clustered sampling. Hence the stochastic models can be used to answer some of the questions introduced above regarding the results of deterministic inversions.

Convection models have advanced, but they are still generic in that they do not claim to know the initial or boundary conditions well enough to be able to simulate the behaviour of the actual Earth. Hence deterministic seismic models are of limited direct use (with the possible exception of planform) to constrain convection simulations, while stochastic seismic models provide directly spectral information which can be used as constraints. The ISC data set contains very large numbers of data ($P \approx 4$ million, $S \approx 0.5$ million) of a very heterogeneous nature (arrivals picked by different observers, on different seismometers, at unique sites); hence the data are well suited for a statistical treatment. Since we are also combining greater numbers of data to evaluate fewer parameters, we should more effectively reduce the effect on the model of any random error left in the data. All of the body wave lower mantle models discussed above have derived the mantle P-wave velocity structure. In this paper we use S-waves to obtain an independent estimate of mantle structure; i.e. its shear velocity structure.

Seismic estimates of the ratio $d \ln (V_s)/d \ln (V_p)$ in the literature have been suggested to be about 2 (Dziewonski & Woodhouse 1987). This value is much higher than that obtained by experiments at room temperature and pressure. Anderson (1987) has suggested that such values in the lower mantle can be explained as the result of the effect of compression on the elastic properties. He has been supported by the molecular dynamics calculations of Agnon & Bukowinski (1989), but this explanation has been questioned by the theoretical calculations of Reynard & Price (1990) and from shock-wave experiments (Duffy & Ahrens 1990). Previous seismic evaluations have involved comparisons of different period data (Dziewonski & Woodhouse 1987), or data with an unknown noise level (Davies & Clayton 1986). Here we compare P and S data sets of the same period range and estimate the noise.

The heart of this stochastic method is evaluating the scatter in traveltime residuals of rays that travel similar paths. Consider two earthquakes occurring at the same location. They should have the same traveltime to all stations. Any difference between the traveltimes to the same station must be a reflection of errors in picking the arrival time, errors in locating the source, errors in the clocks at the stations etc. Consider now an Earth with laterally uniform velocities at all depths except for one layer, where the velocity is lower by a constant amount in one region. Consider two rays that are sufficiently close together, that they both travel equal distances through the anomalous region. They will record the same residual. Next let the rays be spaced sufficiently far apart, such that only one goes through the anomalous region. In this case the two residuals will be different. By discovering at what separation the difference appears we can estimate the size of the anomalous region. From the magnitude of the difference between the residuals we can also evaluate the magnitude of the regional velocity perturbation, assuming that we know the ray length through the region. We can investigate the depth variations of velocity variations in the lower mantle by comparing the scatter in bundles of rays that travel different distances, since rays that travel further penetrate deeper into the Earth. Equally bundles with deep source depths sample only half the upper mantle compared to bundles with shallow source depths, hence this allows us to investigate the depth variations of velocity variations in the upper mantle.

2 EVALUATION OF TRAVELTIME VARIANCE

As discussed above, the heart of the stochastic method involves evaluating the scatter in the traveltime residuals of a bundle of rays. We define the bundle of rays, by means of an unique source and receiver cell on the surface of the Earth. We follow the procedure outlined in GDC, where we divided the earth into a grid of equal area cells. Grids were generated at scales varying from one box for the whole Earth to boxes which are 2° on a side. Rays are binned together if they fall in the same epicentral distance (Δ) range (the distance between source and receiver) and also into the same source depth (Z) bin. A summary ray is a collection (a bundle) of rays that share the same epicentral distance and source depth bins, and also the same receiver and source grid cells. The bins are 4° wide in epicentral distance, extending from 31° to 75°. The bins in source depth are 0–32, 34–60, 60–100, 100–200, 200–400 and 400–700 km. The scale of a grid is defined as the angular radius of a circle on the Earth's surface, which has the same surface area as the average cell in the grid. This measure is denoted by Θ and reported in degrees. We use data collected by the ISC from 1964 to 1986. We select S-wave

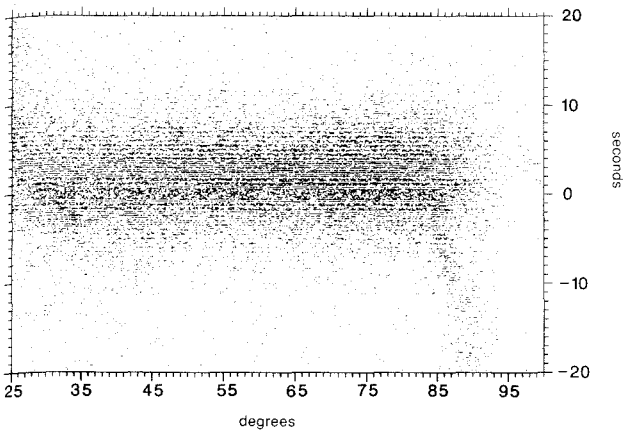

Figure 1. Plot of Jeffreys–Bullen shear wave traveltime residuals as a function of epicentral distance. Notice that their distribution is uniform over teleseismic distances (30° to 100°) with the exception of the vertical streak due to the misidentification of *SKS* as *S* from ≈80° to 90°.

arrivals from events that have at least 50 phases reported because these would be better located. All events located at 0 or 33 km in depth are ignored, since the ISC frequently assigns poorly located events to these depths. We use shear wave arrivals out to only 75° to avoid contamination by *SKS*-waves, whose traveltime curve crosses *S* at about 80°. This contamination can be observed in Fig. 1 as the streak that is directed down and slightly to the right at this distance. Having no data beyond 75° means that we do not sample the lower 450 km of the mantle. Hence, we have no information on the shear velocity structure at the core–mantle boundary (CMB). We selected only shear wave arrivals whose JB traveltime residuals (dt) lie between −9 and +11 s (mean = 1.0 s) to remove obviously incorrect residuals (e.g., 1 min errors, phase mispicks etc.).

Due to problems of triplications, we only use teleseismic phases. Hence, all the rays have similar paths in the upper mantle. Consequently we would have poor radial resolution in the upper mantle. However, earthquakes occur at all depths in the upper mantle and we gain some radial resolution by comparing the scatter in bundles of rays at different source depths. A ray bundle with a shallow source depth has a longer path length on the source-side leg than a bundle with a deep source depth. Hence, it samples shallow heterogeneities that are not sampled by the deep source depth bundle.

To measure the degree of correlation between the traveltime residuals of the component rays comprising a summary ray, we evaluate the variance of the residuals as follows:

$$\sigma_k^2 = \sum_{i=1}^{n_k} (\delta t_i - \overline{\delta t_k})^2/(n_k - 1) \tag{1}$$

where δt_i is the traveltime residual of the ith component ray, $\overline{\delta t_k}$ is the mean residual for the kth summary ray, σ_k^2 is the variance of the kth summary ray, and n_k is the number of component rays in the summary ray.

We only use summary rays which have at least four component rays ($n_k \geq 4$). Then we average the variance of all the summary rays at a given grid size, epicentral distance and source depth bin. To improve the variance estimate we repeated the process on the same grid, but with all the cells displaced one-sixth of their width in longitude. This is done six times and the mean of all six computations is used as the estimate of the variance.

$$\sigma^2(\Theta, \Delta, Z) = \sum_{q=1}^{n_r} \left[\left(\sum_{p=1}^{n_s} \sigma_{p,q}^2/n_s \right) \Big/ n_r \right]. \tag{2}$$

Here p is the summation index over the n_s summary rays that are from the same q, Δ, and Z bin, while q is the summation index over the $n_r = 6$ grid rotations. The variance of an individual summary ray is $\sigma_{p,q}^2$, as evaluated in equation (1). Note that n_s is a function of q, i.e., the total number of summary rays depends on the grid orientation. The above equation may seem complicated, but it is in fact very simple, since it is only an average variance for all the summary rays with at least four component rays, found for all cells and orientations of the grid.

The uncertainty of the estimate is derived from the scatter between the estimates of the six rotated grids. At small scales with grid size less than 5° the data are only included if there are more than 25 summary rays, i.e., $n_s > 25$.

Each epicentral distance and source bin is characterized by a reference ray, which is located at the mean epicentral depth and travels the mean epicentral distance for that data bin. The reference rays are traced through the spherically symmetric Jeffreys–Bullen reference Earth model, which is a reasonable model for locating ray paths at teleseismic epicentral distances. From Fig. 1 we see that the residuals are symmetrically distributed about their mean for teleseismic rays ($\Delta > 30°$) demonstrating that there is little bias in the reference model. In this study we use 66 reference rays, 11 epicentral distance bins, each with six epicentral depth bins.

The raw data $\sigma^2(\Theta, \Delta, Z)$ are presented in Figs 2 and 3. Half the data at all the scales are presented in Fig. 2. The scale axis is logarithmic, allowing the data at all scales to be reasonably viewed. Interestingly the data approximately describe a straight line, suggesting a possible relationship for extrapolating variance to small scales. In Fig. 3 we present data for two selected depth bins close to the origin with a linear scale axis. Notice that these data also describe a reasonable straight line. We also observe a general trend of decreasing variance at decreasing scale as would be expected, since the rays are travelling ever more similar paths. Note the knee in the data that occurs at a scale of 5°–10°. For uniform sampling this would imply that the signal-averaged scalelength is of the same order. Since the sampling is clustered the effective grid size is less than the actual grid size, hence the actual scalelength is found to be shorter. We also observe that the variance decreases with increasing source depth. This is as expected since path length decreases as source depth is increased. The variance varies weakly with epicentral distance. This implies that the strength of heterogeneity must decrease with depth, since if the Earth had constant heterogeneity throughout then the variance would increase linearly with path length [see equation (11) below]. Finally, we observe that the variance approaches a finite value at small scales. We interpret this intercept as the random error variance in the data. In summary, the data curves follow our intuition in their

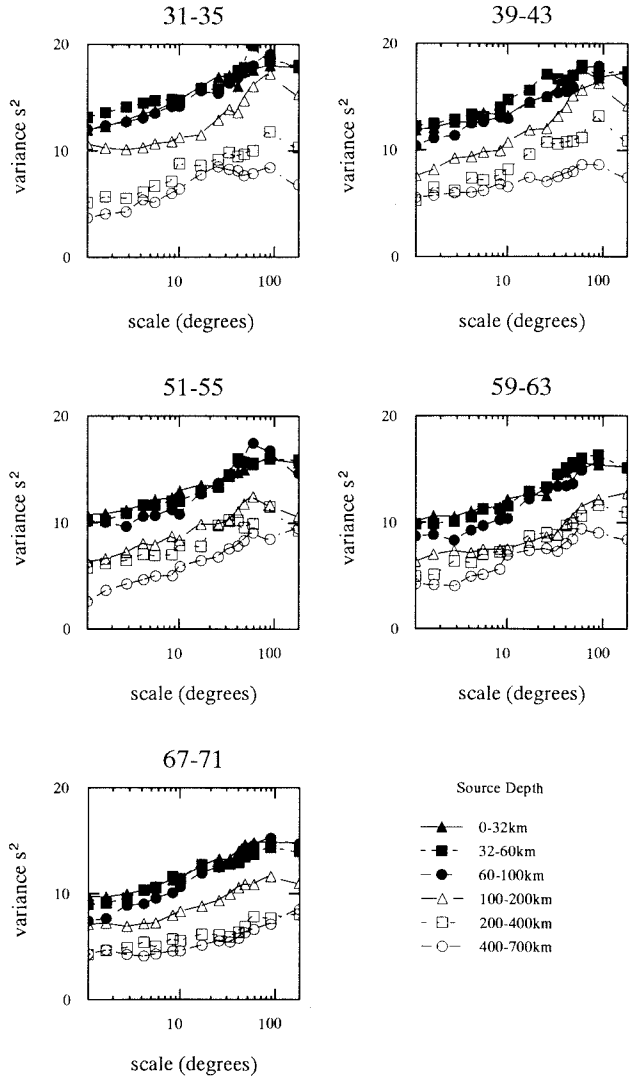

Figure 2. Variance of a selection of reference rays plotted versus grid size. Note the logarithmic scale for the grid size axis. The variance decreases regularly as scale of grid decreases, and the data from the deepest depth bins (open symbols) are lower than the shallower data; e.g. the filled squares, circles and triangles.

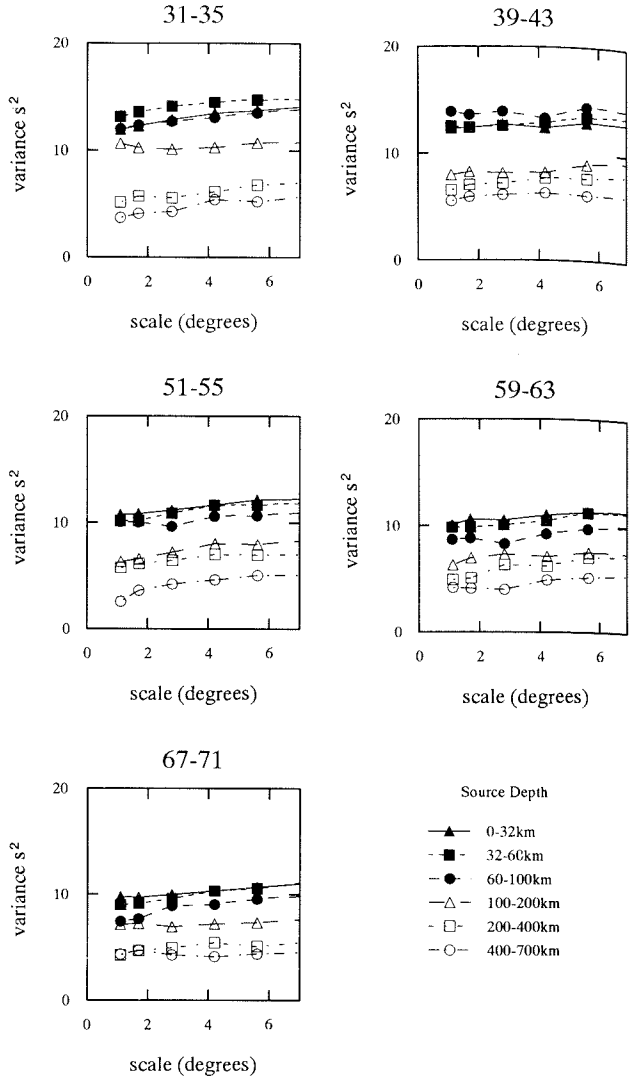

Figure 3. Variance of a selection of data plotted versus the scale of grid size, for small grid sizes. Note scale axis is linear.

behaviour and are reasonably coherent from curve to curve. This gives us hope that the data can be reasonably modelled in terms of Earth structure.

We estimate the intercept by extrapolating all the data for a given depth bin of a scale smaller than 5°, assuming that the decay is linear and that the slope is fixed as $b = B/\cos(i)$, where i is the angle of incidence at the receiver. The linear decay is supported by the fact that the data at the smallest scales are well described by a straight line, as seen in Fig. 3. Assuming that most of the small-scale signal is located at shallow depths close to the receiver, we weight the slope by $1/\cos(i)$, which is proportional to the path length near the surface. The different source depth bins require different intercept estimates. This suggests that the incoherent signal is not solely due to different estimates of the smallest scale signal for different depth bins, but reflects a real decrease in noise with increasing source depth.

In Fig. 4 we illustrate the intercept estimates for all the reference rays. We find that the estimates decrease with source depth and are relatively constant as a function of epicentral distance. The average variance for the intercepts for sources shallower than 100 km is $\approx 8\,s^2$ as compared to a maximum total signal at the largest scales of $\approx 18\,s^2$. The uncertainty in estimating the intercept is high. Hence, the residual data arising from subtracting the intercepts from the original data have a proportionately higher uncertainty. The extrapolation of the intercept could be done logarithmically rather than linearly. The straight line that the data describe at larger scales in Fig. 2 supports this choice. The logarithmic intercepts are shown in Fig. 4(b). We have used both sets of intercepts estimates in our inversion of the residual variance for the statistics of Earth structure. We observed that the differences in the results are restricted to small-scale structure. The residual variance data that result from the subtraction of linear intercept estimates (Fig. 4a) are shown in Fig. 5. It is the residual variance data, following the subtraction of linear intercept estimates, that

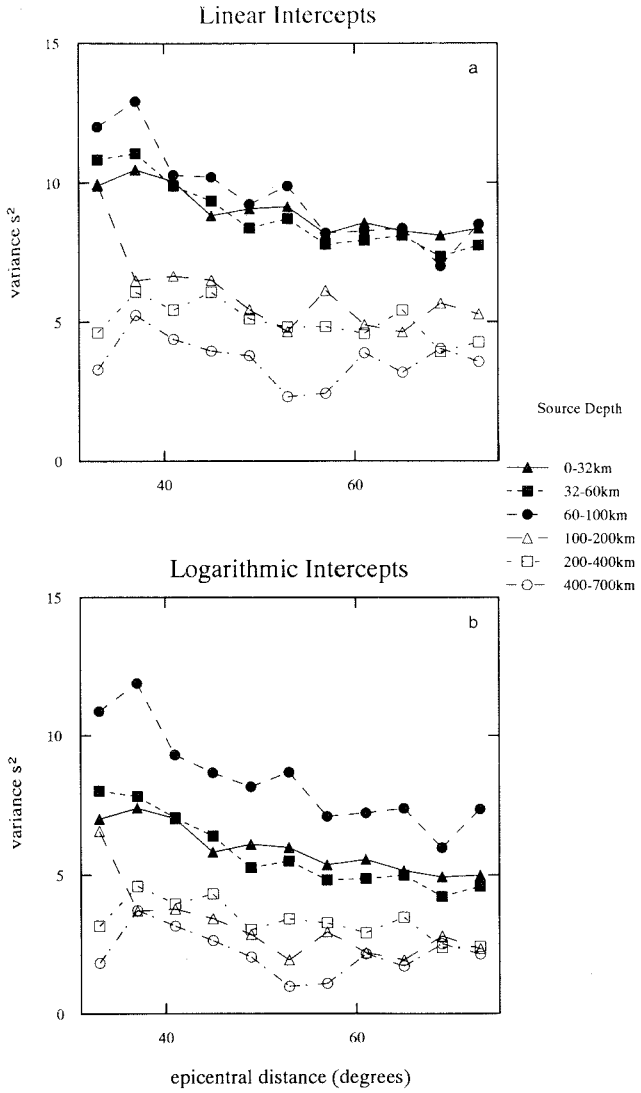

Figure 4. Intercept estimates versus epicentral distance. (a) Linear extrapolation. (b) Logarithmic extrapolation.

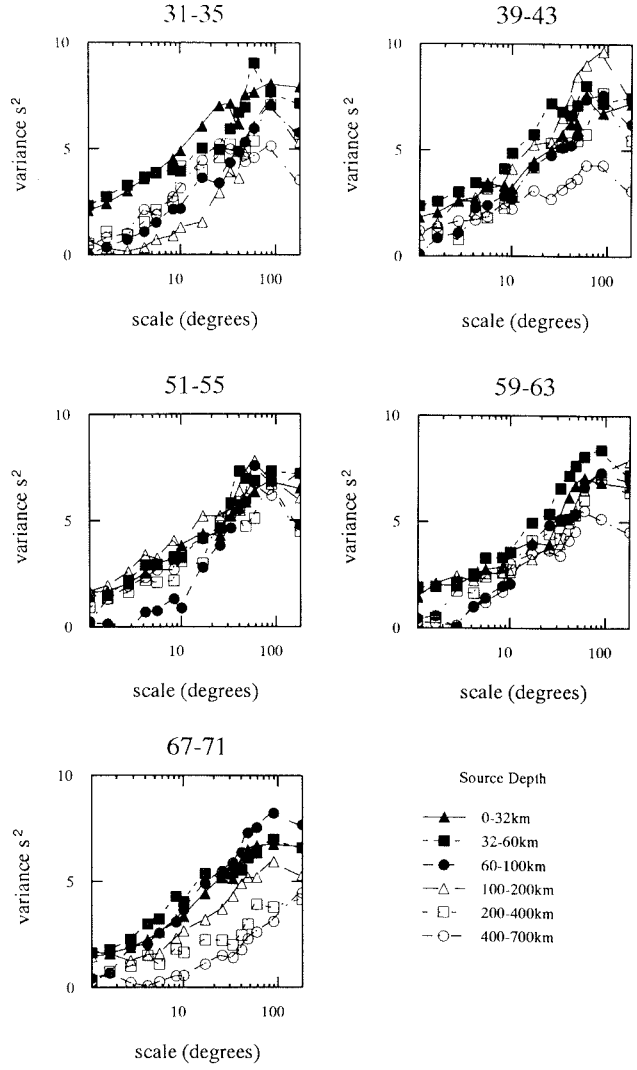

Figure 5. Data after removal of the linearly extrapolated intercepts of Fig. 4(a).

are used to constrain the spectrum of mantle heterogeneity via a linearized inversion procedure. Note that the intercept estimates include errors arising from incorrect hypocentral parameters, hence their contamination of estimates of structure is limited. Before developing the inversion procedure though we must derive a forward model to relate the variance to the lateral variations of velocity structure; this is outlined next.

3 LINEAR MODEL RELATING TRAVELTIME VARIANCE TO SPECTRA OF MANTLE STRUCTURE

Using ray theory we develop a linear theory to relate the variance of the traveltime residuals of the different bins of rays at different grid sizes to the spectra of the seismic heterogeneity as a function of depth through the mantle. We do this in two stages; first we relate the variance of the summary rays to the autocovariance of traveltime residuals, and second we relate the autocovariance of the traveltime residuals to the autocovariance of the slowness perturbations in the medium. Since the details of the derivation of the forward model have been given previously in GDC, we shall only outline the derivation.

The relationship between the variance of the summary rays $\sigma^2(\Theta, \lambda, Z)$ and the autocovariance of the traveltime residuals $[T(\rho)$, where ρ is the separation of the endpoints] is geometrical, and is represented by a weighting function $[w(\Theta, \lambda)]$. To account for non-uniform sampling we also introduce the autocovariance of the sampling function $[B(\Theta, \lambda)]$. By assuming that the sampling is independent of the structure, we can separate w and B, and hence we get

$$\sigma^2(a) = \int_0^{2a} w(a, \rho)[T(0) - T(\rho)]B(a, \rho)\, d\rho. \tag{3}$$

By using ray theory, and assuming that our reference model is linearly close to the real Earth, we can apply Fermat's principle and linearly relate the traveltime residual linearly to the slowness perturbations along the ray path

$$\delta t(\eta) = \int_{\text{ray path}} \delta U(x)\, dS \tag{4}$$

where δt is the traveltime residual, U is the slowness (1/velocity) and S is the path length. Hence the autocovariance of the traveltime residuals, $T(\rho)$, can be related to the autocovariance of the medium, $R(x)$ (which is assumed to be a function of separation only and not of direction, i.e. we assume isotropy) as follows:

$$T(\rho) = E[\delta t(\eta_1)\delta t(\eta_2)] \tag{5}$$

where $\rho = |\eta_1 - \eta_2|$ is the separation of the endpoints, and E is the expectation operator. By substitution of Fermat's equation we get

$$T(\rho) = \int_{\text{ray1}} \int_{\text{ray2}} E[\delta U(x_1)\delta U(x_2)]\, dS_1\, dS_2, \tag{6}$$

$$T(\rho) = \int_{\text{ray1}} \int_{\text{ray2}} R(|x_1 - x_2|)\, dS_1\, dS_2. \tag{7}$$

To evaluate one of the integrals we approximate the ray geometry by two parallel rays and assume that the radius of curvature of the rays is greater than the scalelength of the anomalies; then it can be shown that

$$T(\rho) = dC \int_{\text{ray}} x_{1/2} R(\rho)\, dS, \tag{8}$$

where C is a constant of order 2; its exact value depends upon the medium autocorrelation function but is exactly $(\pi/\ln 2)^{1/2}$ if it is Gaussian, while $x_{1/2}$ is the half-width of $R(\rho)$, i.e. $R(x_{1/2}) = 0.5R(0)$.

For the spherical Earth we have parametrized the slowness perturbations as follows:

$$\delta U(r, \theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} A_{lm}(r) Y_{lm}(\theta, \phi), \tag{9}$$

where $Y_{lm}(\theta, \phi)$ are the fully normalized spherical harmonics, θ is the colatitude and ϕ is the longitude. A_{lm} are the coefficients of the appropriate spherical harmonic, l is harmonic degree and m is harmonic order. Hence the medium autocorrelation function can be written as

$$R(\lambda, r) \approx \frac{1}{4\pi} \sum_{l=0}^{\infty} (2l+1) Q_l(r) P_l(\cos\lambda) \tag{10}$$

where

$$Q_l = \frac{1}{2l+1} \sum_{m=-l}^{l} A_{lm} A_{lm}^*,$$

where * denotes complex conjugate, λ is the angular separation of the two points and P_l is the Legendre function of degree l.

Hence by combining the equations above we find the equation relating residual summary ray variance, σ^2, (with the estimated intercepts removed) at scale Θ, to the power spectrum, Q_l, of the structure:

$$\sigma^2(\Theta) = \frac{C}{4\pi} \sum_{l=0}^{\infty} \int_0^{2\Theta} \int_{\text{ray}} w(\Theta, \lambda) B(\Theta, \lambda)(2l+1) x_{1/2}(r) Q_l(r) \times [1 - P_l(\cos\lambda)]\, dS\, d\lambda. \tag{11}$$

The above equation can be cast in matrix form as

$$\mathbf{D} = \mathbf{G}\mathbf{X}\mathbf{F}^{\mathrm{T}} \tag{12}$$

where

$$D_{ij} = \sigma_i^2(\Theta_j), \qquad G_{ik} = \frac{dS_i}{dr} \Delta r_k,$$

$$F_{jl} = \frac{C}{4\pi} \int_0^{2\Theta_j} w(\Theta_j, \lambda) B(\Theta_j, \lambda)[1 - P_l(\cos\lambda)]\, d\lambda,$$

$$X_{kl} = (2l+1) x_{1/2}(r_k) Q_l(r_k),$$

which states that the data, $\mathbf{D}$, can be related to the model, $\mathbf{X}$, by two kernel matrices. One is G_{ik}, which relates the variance of the ith reference ray to the value of the model at the kth depth. The second if F_{jl}, which relates the variance at scale Θ_j to the model's lth harmonic degree.

An intuitive explanation for why the variance of traveltimes is proportional to the product of the variance of slowness variations and a scalelength of the variations integrated along the ray length can be developed by analogy to the random walk of a drunkard. Consider a 1-D medium consisting of equal length segments in which the slowness perturbations are of constant magnitude, but random in sign. A ray path through such a medium can be thought of as a random walk. We can think of a ray as stepping from one anomaly to the next with the cumulative traveltime residual increasing or decreasing by a fixed amount. This is equivalent to the 1-D random walk, where the drunk can go to the right (positive residual) or to the left (negative residual). The most probable result is a zero residual (no digression), but the variance of the residuals is proportional to the number of steps times the square of the residual of a single step, as in a random walk. The residual of a single step is proportional to the product of the slowness perturbation and its width (scalelength). The number of steps is proportional to the ray length divided by the perturbation width. Hence, the expected traveltime variance is proportional to the ray length, the scalelength and the square of the slowness perturbation.

4 INVERSION

In the previous section we have developed a linear formulation for the forward problem describing the effect of lateral heterogeneity, X_{kl} (the product of the half-width of the correlation function of slowness variations and the mean of the square of the slowness perturbations, hereafter referred to as the scattering power) on the variance of summary rays, D_{ij}. Hence we can invert equation (11) using any one of a number of linear inversion techniques. We chose to use a damped least-squares method, which gives

$$\mathbf{X} = (\mathbf{G}\mathbf{G}^{\mathrm{T}} + \alpha^2\mathbf{I})^{-1}\mathbf{G}^{\mathrm{T}}\mathbf{D}\mathbf{F}(\mathbf{F}\mathbf{F}^{\mathrm{T}} + \beta^2\mathbf{I})^{-1} \tag{13}$$

which can also be written as

$$\mathbf{X} = \mathbf{G}^{\mathrm{T}}(\mathbf{G}\mathbf{G}^{\mathrm{T}} + \alpha^2\mathbf{I})^{-1}\mathbf{D}(\mathbf{F}\mathbf{F}^{\mathrm{T}} + \beta^2\mathbf{I})^{-1}\mathbf{F}. \tag{14}$$

This is a 2-D inverse problem with two rather than one kernel matrix and a model matrix rather than a model vector. The damping parameters are α and β.

We singular value decompose the kernel matrices, $\mathbf{G}$ and $\mathbf{F}$,

$$\mathbf{G} = \mathbf{U}\Lambda\mathbf{V}^{\mathrm{T}}, \tag{15}$$

and

$$\mathbf{F} = \mathbf{R\Omega S}^{\mathrm{T}}, \tag{16}$$

where $\mathbf{U}, \mathbf{V}, \mathbf{R}, \mathbf{S}$ are the matrices of the left and right eigenvectors corresponding to non-zero singular values and $\mathbf{\Lambda}$, $\mathbf{\Omega}$ are the diagonal matrices of the singular values. Using equations (15) and (16), (14) can be rewritten as

$$\mathbf{X} = \mathbf{V\Lambda}(\mathbf{\Lambda}^2 + \alpha^2\mathbf{I})^{-1}\mathbf{U}^{\mathrm{T}}\mathbf{DR}(\mathbf{\Omega}^2 + \beta^2\mathbf{I})^{-1}\mathbf{\Omega S}^{\mathrm{T}} \tag{17}$$

where we have used that $\mathbf{R}^{\mathrm{T}}\mathbf{R} = \mathbf{U}^{\mathrm{T}}\mathbf{U} = \mathbf{V}^{\mathrm{T}}\mathbf{V} = \mathbf{S}^{\mathrm{T}}\mathbf{S} = \mathbf{I}$, and a similar transformation to the one that is used to relate equation (13) to equation (14). Notice that the only matrices which need to be inverted are diagonal, and hence, can be inverted trivially. This reduces the cost of computations that require repeated inversion, such as a search for suitable damping parameters.

Least-squares inversion minimizes the prediction error, i.e., the sum of the squares of the differences between the data and the data as predicted by the model. Damped least-squares inversion minimizes a linear sum of the prediction error and the L2 norm of the model (in this case we can think of our model matrix disassembled into a single long vector, in order to define an L2 norm). The damping parameters determine the relative importance of data prediction and model minimization. We selected the damping parameters such that model error was minimized, while depth and spectral resolution were maximized. We investigated a number of different linear combinations of the above model measures and found that the trade-off surface changed little. An example of a trade-off surface is shown in Fig. 6. Solutions from within its well (defined by the smallest contour) were investigated and found to be largely insensitive to the exact choice of damping. The small cross with vertical and horizontal bars shows the damping parameters chosen for the presented solution. The cross with diagonal bars shows the damping parameters used in an alternate solution to illustrate which features of the solution are robust. This choice of damping effectively limits the solution to nine spectral degrees of freedom and 5–6 radial degrees of freedom.

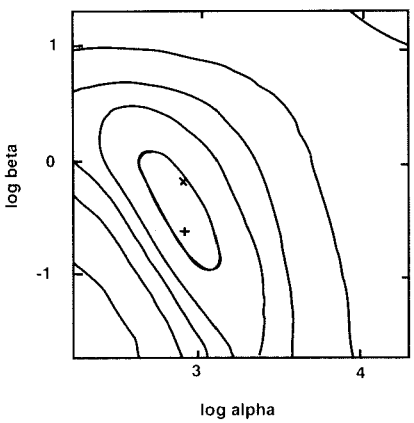

Figure 6. Trade-off surface for inversion. The contoured surface illustrates how a sum of the model error and the resolution lengths of both the depth and spectral matrices varies with the depth (alpha α) and spectral (beta β) damping parameters. The cross with the horizontal and vertical bars shows the position of the favoured solution (Fig. 9) while the diagonal cross shows the damping parameters for the alternate solution (Fig. 10); both are obviously close to the minima. Since the normalizations and weighting of the different components are slightly arbitrary it is obviously meaningless to search for an exact minima. Solutions from this region of parameter space were similar. An exception is shown in Fig. 10.

Notice that damped least-squares modelling leads to global damping, i.e. the same value of the damping parameters is used in deriving the whole model. We have poor depth resolution in the upper mantle (since no rays bottom there), while we have relatively large error variance in the lower mantle (since the model is small there). Thus, we would prefer to tune the trade-off differently in these two parts of the model. This problem can be partly alleviated by a suitable choice of parametrization. Given the large uncertainties in the finely parametrized model STP1 of GDC we use a coarse depth parametrization in our model, similar to that of model STP2 of GDC. The depth boundaries in the model, giving the radial parametrization are at 0, 61, 300, 540, 670, 970, 1470 and 2450 km depth, i.e. seven depth layers.

The spectral parametrization was chosen to be parabolic, i.e., we combine a number of harmonic degrees to give one spectral model parameter, with the number of harmonics summed into a single model parameter increasing approximately parabolically with the harmonic degree of the mid-point. We invert for 50 such parameters, the first 11 representing single harmonics, with the following parameters representing more than one harmonic degree, with the number increasing parabolically until the last parameter represents 40 harmonic degrees, from 460 to 500. Notice that the smallest grid that we use has a scale of 2°. Hence, our data do not constrain any power present below this scale.

We define a simple measure of resolution length in both depth and spectra for display purposes as follows:

$$L_j = \Delta M_j \times \left(\frac{1}{1 - \left[(1 - Res_j)^2 + \sum_{i=0 \text{ and } i \neq j}^{n} Res_i^2 \right]} \right) \tag{18}$$

where L_j is the simple measure of the resolution length of the jth element, ΔM_j is the size of the element, i.e. when considering spatial resolution this is the thickness of the corresponding layer in km, while when considering spectral resolution it is the number of harmonic degrees combined in the spectral parameter, and finally Res_i is the sum of all the elements of the ith column of the appropriate resolution matrix as defined below.

The spectral and spatial (radial) resolution matrices are

$$\text{Spectral resolution} = \mathbf{S\Omega}^2(\mathbf{\Omega}^2 + \beta^2\mathbf{I})^{-1}\mathbf{S}^{\mathrm{T}} \tag{19}$$

and

$$\text{Radial resolution} = \mathbf{U\Omega}^2(\mathbf{\Omega}^2 + \alpha^2\mathbf{I})^{-1}\mathbf{U}^{\mathrm{T}}. \tag{20}$$

The spectral resolution (see Fig. 7a) behaves nearly linearly with harmonic degree. This is due to the uneven sampling of the scale axis. The depth resolution (see Fig. 7b) is nearly identical to the radial parametrization in the lower mantle. The poorest resolution occurs just above the 670 km discontinuity and in the shallowest layer. The resolution is relatively poor in the upper mantle, but

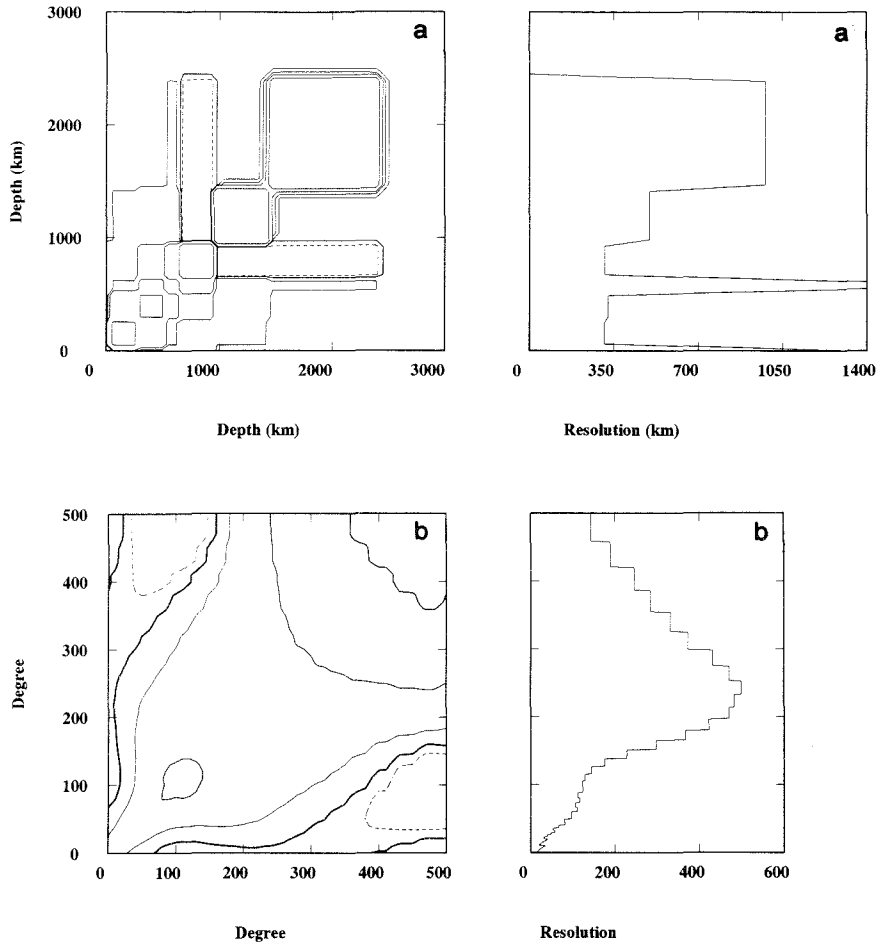

Figure 7. Radial (a) and spectral (b) resolution matrices, on the left, with the simpler measure of resolutions as given by equations (19) and (20) on the right. Note both radial figures (a) have the same vertical depth axis, while both of the spectral (b) figures have the same vertical harmonic degree axis.

because of the coarse parametrization of the lower mantle, each model parameter is nearly perfectly resolved.

In evaluating the **F** matrix, the clustering of stations and events was taken into account, i.e. $B(\Theta, \lambda)$ was evaluated. This was done by evaluating the frequency of event pairs and station pairs of a given separation, which is proportional to $B(\Theta, \lambda)w(\Theta, \lambda)$. Hence, we can derive B by dividing the histograms by the known kernel functions, $w(\Theta, \lambda)$. This was repeated over different scales. Much like GDC we found that the functions, B, are reasonably described by power laws of separation (see Fig. 8) with the exponent decreasing as scale increases (at vanishing scales the exponent must be infinite). The variation in the exponent with scale was found to be well fit by a linear relation. Since this is computationally very expensive the distribution was evaluated for the data of a single reference ray (51°–55° and 1–32 km), and was assumed to hold for all other reference rays. The distribution for the S-waves was found to be a weaker function of scale than for the P-waves (GDC). This is not unexpected since the P-wave data set is approximately 5 to 6 times larger, while the pattern of seismicity and station distribution is little changed between the two data sets.

Using kernels based upon a uniform sampling distribution would lead to overestimates of correlation length, and

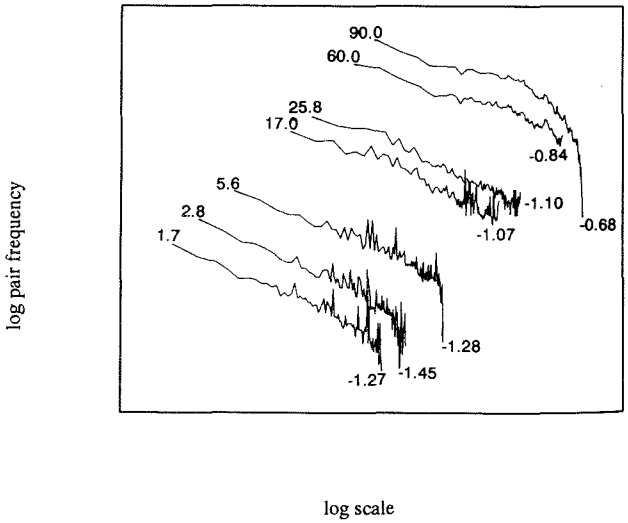

Figure 8. Plot of the normalized log pair frequency as a function of separation of the stations and events for different scales. This is evaluated for the data in the epicentral distance bin from 51° to 55°, and the shallowest epicentral depth bin. The scales are labelled in degrees above the curves while the negative numbers below represent the least-squares estimates of the linear gradients of the curves.

hence, underestimates of slowness perturbations. The S-waves show a marked deviation from a power-law relation at the largest scales, see the curve for scale 90° in Fig. 8. The curve shows a large deficiency of pairs at large separations. Therefore, we have few data at large scale, which are related to large separations. Hence, sensitivity to the largest scale heterogeneities is limited. This follows from the fact that no major zones of seismicity are antipodal; equally there are no antipodal continents.

5 RESULTS

The results of our inversion of residual variance for the statistics of Earth structure are presented in Figs 9(a), (b), (c), (d), (e) and (f). In Fig. 9(a) we present the total scattering power (the product of half-width and the slowness variance). It is highest in the upper mantle and decreases with increasing depth. There are two exceptions to this behaviour. The surface value is low, and there is an

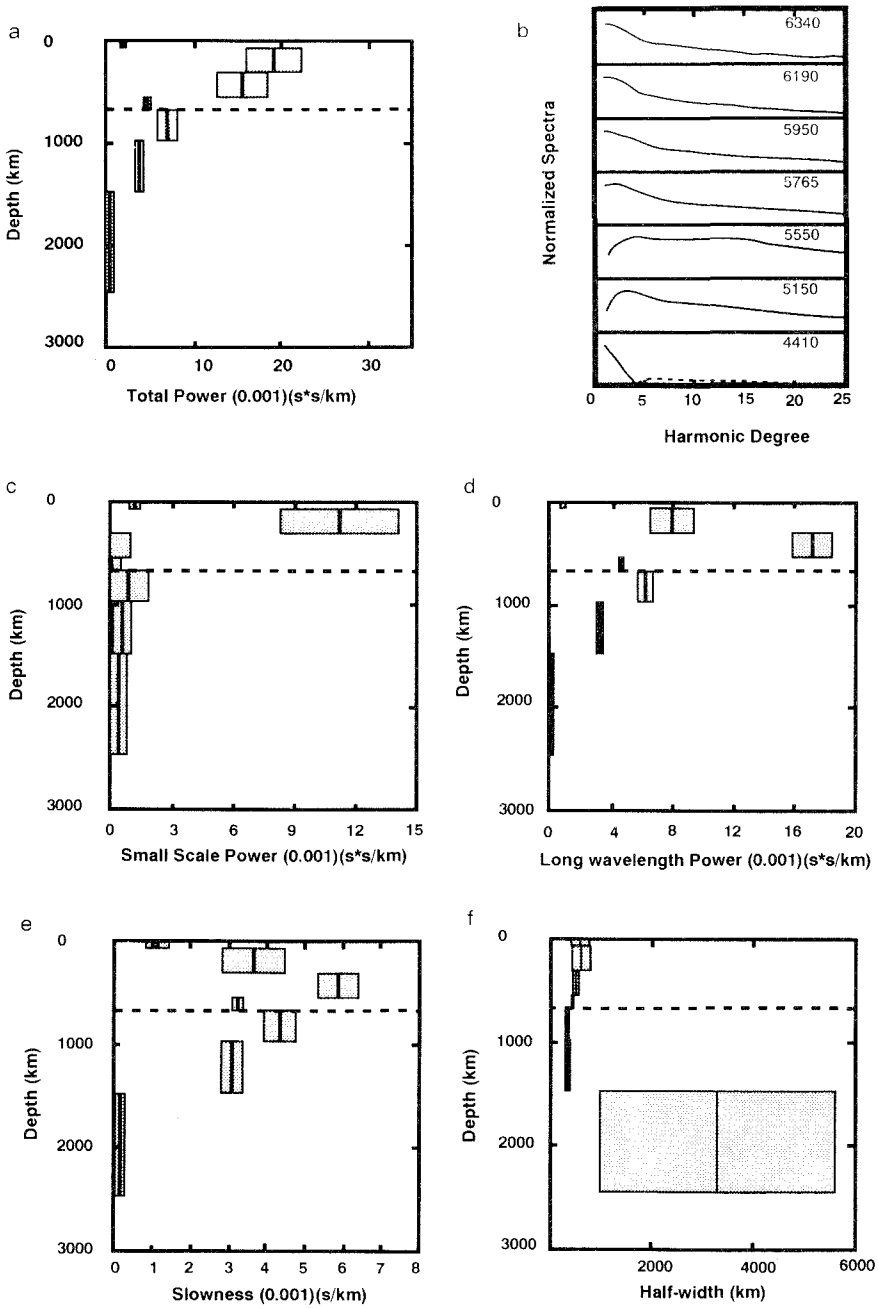

Figure 9. Results. (a) Total scattering power (the product of half-width and the square of the mean slowness); the width of the box represents one standard deviation error. (b) Normalized spectra; the dashed line represents negative power and the number in each box represents the mean radius of that model bin. (c) Short-scale scattering power ($l > 50$); notice power concentrated in uppermost mantle. (d) Large-scale scattering power ($l < 50$). Power is concentrated in the upper mantle and decays with depth in the lower mantle. Notice more power below than above 670 km discontinuity. (e) Slowness perturbations away from reference model. Notice pattern similar to large-scale power. (f) Half-width estimates of large-scale correlation function, around 300–600 km for uppermost mantle. Poorly constrained in lower mantle.

increase in power across the 670 km discontinuity. Both exceptions occur in regions of poor depth resolution.

In Fig. 9(b) we present the lowest 25 harmonic degrees of the spectra for all seven depth bins. They are normalized to unit height. Dashed curves signify negative power, an unphysical result. The negative values are small. Their magnitude is probably a fair indication of the real uncertainty in our solution (including random and systematic data errors and forward modelling errors).

Since the spectral resolution deteriorates with increasing harmonic degree, we decided to split the spectrum into two parts; a long-wavelength part and a short-wavelength part. We believe that we can make a reasonable estimate of the half-width of the long-wavelength part, whereas the estimate derived from the whole spectrum would probably not be as meaningful. The cut-off between the two parts was arbitrarily chosen at harmonic degree 50.

In Fig. 9(c) we present the short-scale power, the part of the total power that is due to structure of harmonic degree >50. Nearly all this power is concentrated in the second depth bin (60–300 km depth). There is very little small-scale power in the lower mantle. Some of the values (i.e., the third depth bin) are negative. In Fig. 9(d) we present the integral long-scale power, due to structure of harmonic degree <50. It is largest in the third layer (i.e., from a depth of 300–540 km), and decays through the lower mantle. The layer just below the 670 km discontinuity has more power than the layer above. The negative small-scale power and the huge large-scale power in the third layer suggests that the spectrum for this layer is extremely focused in the low harmonic degrees. A comparison with the other layers of both the *P*-wave study (GDC) and the present *S*-wave study suggests that this is aberrant.

In Fig. 9(f) we present the half-width, which is estimated by evaluating the scale at which the large-scale correlation function has fallen to half its peak value. By 'large-scale correlation function', we mean a correlation function that is constructed from only the large-scale portion of the power spectrum. By dividing this estimate of the half-width into the large-scale power (Fig. 9d) we obtain the estimates of root mean square RMS slowness, shown in Fig. 9(e).

The main difference between solutions with alternate values of the damping parameter is that the long-wavelength half-width in the lowermost mantle is now 400 km as opposed to 3000 km; otherwise the results are very similar. Hence, the large half-width in the lower mantle in Fig. 10 is not a robust feature. It is probable that the long-wavelength half-width in the lowermost mantle is close to the better constrained estimate of 1200 km obtained by GDC from *P*-wave data.

In Fig. 11 we compare the data predictions of the model and the original residual variance data. The predicted data lie within the error bars of the data, but in Fig. 11(b) they appear to be offset by a constant, i.e., a different intercept estimate could lead to a better data fit.

The variance reduction by the model in Fig. 9 is 94 per cent. The chi squared value (χ^2) is approximately equal to the number of degrees of freedom (945) [number of data ($66 \times 15 = 990$), effective number of model parameters 45 (five degrees of freedom in depth and nine spectral degrees of freedom)]. Hence, the model fit is significant and highly unlikely to be the result of a random distribution of data. Note that there is a potential confusion in comparing the variance reduction of this model with that of deterministic models since our initial data are already variances of traveltimes.

A significant part of the original signal is taken up in the intercept estimates and does not enter into the structural model. The intercepts and model are not evaluated simultaneously but sequentially. This leads to a simpler and more stable inversion scheme. Different schemes of estimating the intercepts can lead to large differences in the estimates. To see how these differences affect the structural model, we inverted a data set where the intercepts were evaluated using logarithmic rather than linear extrapolation. These intercept estimates are lower and the variance left to be inverted is higher (average variance is $6.5s^2$ as against $4.4s^2$). The variance reduction (95 per cent) and χ^2 value are similar to the model derived from data using linear intercept estimates. The primary effect on the model is an

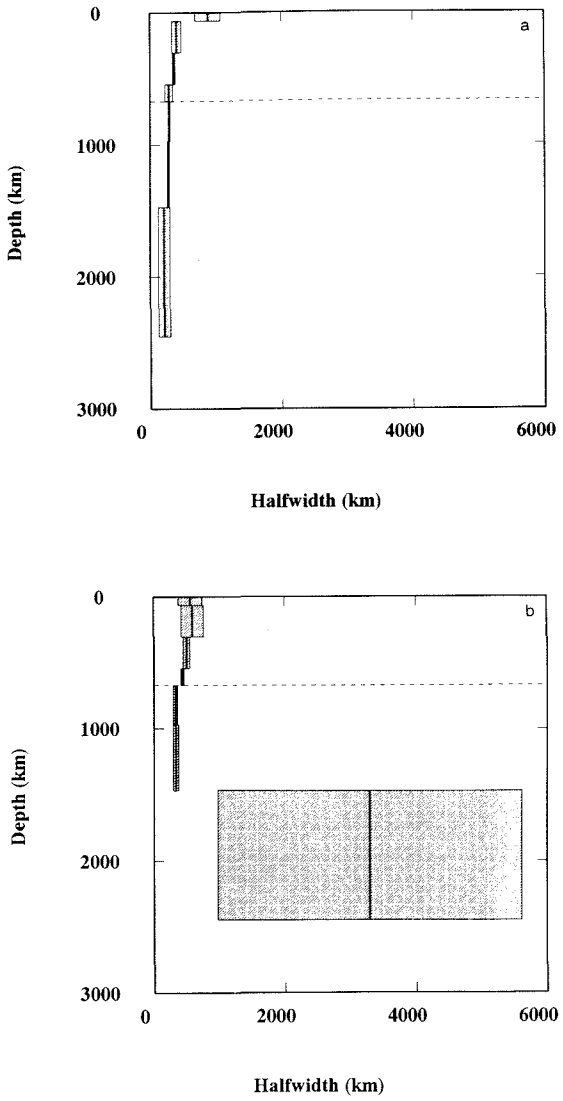

Figure 10. (a) Half-width of model resulting from the inversion of data with different damping parameters (the diagonal cross of trade-off surface in Fig. 5). (b) Half-width of original results. Notice that the value obtained for the half-width of the lowermost mantle is not robust.

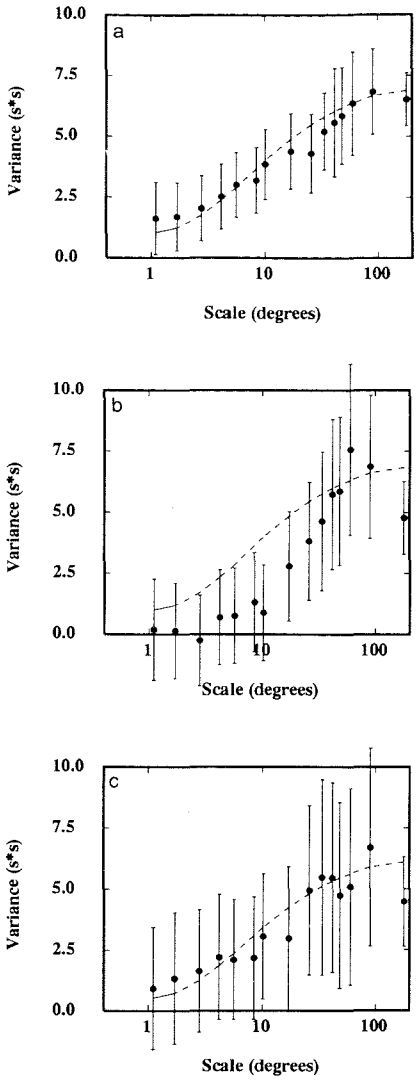

Figure 11. Comparison of data predicted by model to actual data. All with intercepts removed. They are epicentral distance bin 51°–55°; (a) epicentral depth bin 0–32 km, (b) epicentral depth bin 60–100 km and (c) epicentral depth bin 200–400 km.

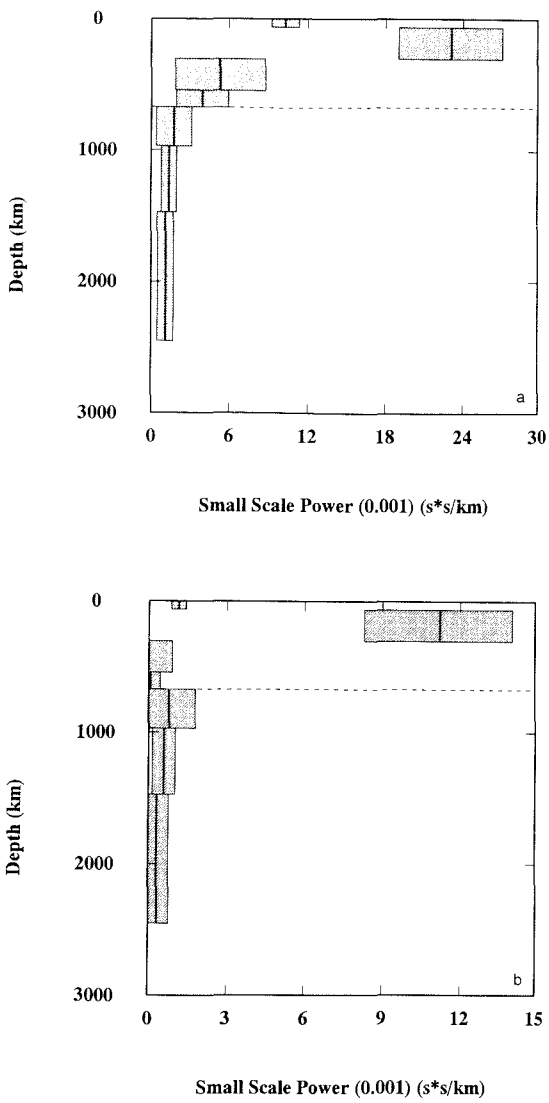

Figure 12. (a) Small-scale scattering power of data with logarithmically extrapolated intercepts removed. (b) Small-scale power of original results, which have linearly extrapolated estimates of intercepts removed from data. Notice that there is more small-scale power when the logarithmically extrapolated intercepts are used; notice the different scale! Given that one can make reasonable arguments for both extrapolations, the difference between the two is a measure of the impact of the uncertainty.

increase in short-scale power in the depth range from 50 to 300 km (see Fig. 12). The short-scale power in the lower mantle is also slightly increased. The large-scale power is virtually unchanged (see Fig. 13), except for the half-width in the poorly resolved shallowest depth bin. These differences demonstrate the trade-off between intercept estimates and small-scale power due to the lack of data at small scales.

The model error variance is low throughout the model. We note, however, that it is possible that the data contain errors which are spatially coherent, and thus depend on scalelength. These errors would not contribute to the intercept estimates, which are measures of the incoherent noise, and are indistinguishable from the residual variance, which we interpret as structural. The only way to confirm the absence of appreciable coherent noise (systematic error) is to obtain an equivalent model using an independent data set. Our results will be shown to be consistent with previous work, suggesting that systematic errors are not a severe problem; this is reinforced by the limited amount of unphysical negative power (Fig. 9b). Negative power arises from trying to fit data where the variance decreases as scale- or path length increases.

Because of the large damping in the upper mantle due to the poor resolution we attempted to model the data using only shallow structure, to see whether we were underestimating upper mantle structure and whether the data required structure in the lower mantle. Good data fits were obtained provided that at least two layers were included (91 per cent variance reduction, chi squared slightly greater than the number of degrees of freedom). The small-scale structure was unchanged, while the large-scale structure increased by up to 50 per cent and RMS slowness by up to 20 per cent. Thus, little is sacrificed in data fit as we simplify the model greatly and restrict it to the upper mantle. This

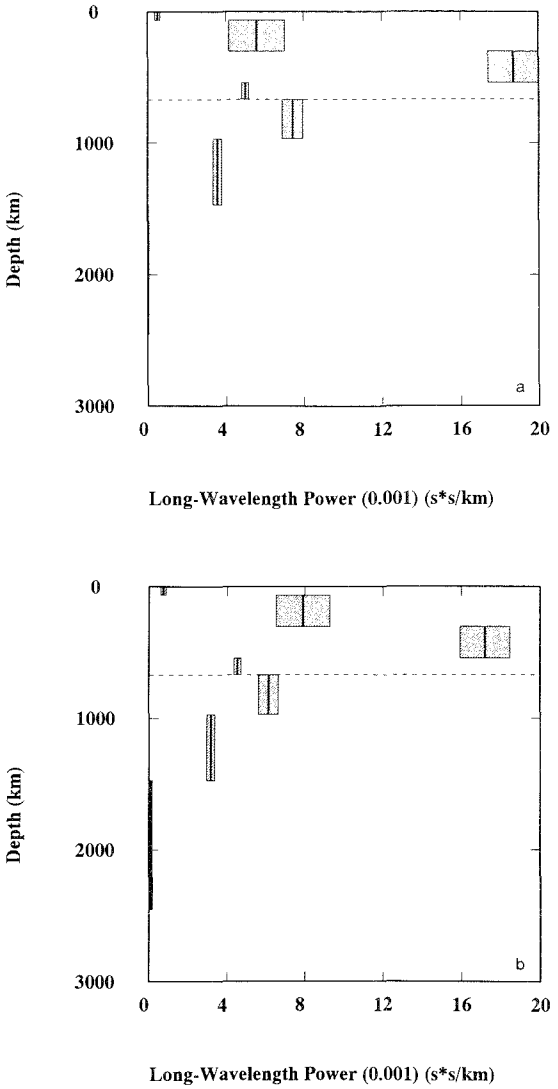

Figure 13. (a) Large-scale scattering power for model resulting from inversion of data with the dip at large grid scales removed by lifting the data curves at the largest scales such that they are flat. (b) Large-scale power from original data. The difference is probably an indication of the uncertainty arising from the imperfect assumption of parallel rays at the largest scales.

suggests the possibility of overestimating the degree of heterogeneity in the lower mantle by power leakage as a result of poor vertical resolution in the upper mantle. Note though that the RMS slowness in the lower mantle is already low, within two standard deviations of zero, and that this test did favour, if only slightly structure in the lower mantle from the increased value of χ^2. This does illustrate the difficulties facing traveltime inversions for lateral heterogeneity of the lower mantle, the core–mantle boundary and the core.

Before discussing implications arising from these results, we should like to remind the reader of their limitations. First, the theory is based on ray theory which is valid only in the high-frequency limit while we use finite-frequency data. Hence, we have intrinsic averaging and a minimum resolution, which generally deteriorates with depth. We would have to consider diffraction to obtain the whole structure. However, the sparseness of data at the small separations is probably a more severe limitation to our resolution at small scales.

Second, the assumption that rays in a summary ray can be modelled as parallel is questionable at large scale. Only an expensive synthetic study of a wide range of models using the actual ray distribution could verify its validity. If it is a problem, it should be restricted to the lowest harmonics. We also have incomplete sampling, which could manifest itself in the correlation of structure and sampling. In particular, most of the deep seismic sources are in a few subduction zones, which are known to have a unique seismic signature, i.e. seismically fast. Hence, our results may be biased towards these parts of the Earth.

Caution is also suggested by certain features in the data that are unexplainable by the theory of GDC. At the largest scale the variance frequently decreases. To explain this, the theory requires unphysical negative model power at large scales. We suggest that this may be explained, as mentioned above, by deviations in the ray geometry from parallel as assumed by the theory, at scales larger than the maximum epicentral distance. A second possible explanation is that the sampling is biased, e.g., more towards oceanic structure as scales increase. A third feasible, but in our opinion unlikely explanation, is that the dip at large scale is a reflection of a very large-scale (harmonic degree 2 or 3) periodic convection pattern. We studied the effect of this feature by artificially increasing the large-scale data, such that the variance became flat. The resulting model is very similar to the preferred model. There are some differences, primarily in the long-wavelength power illustrated in Fig. 13, especially near the 670 km discontinuity. The increase in large-scale power is expected. It gives a quantitative indication of the potential effect of biased sampling and imperfect ray geometry on our results, i.e., 20 per cent for the depths where the resolution is good and the model appreciable.

The low level of power in the shallowest layer in our model may be artificial. It may be caused by artificially large intercept estimates for the shallowest data depth-bin, which results in low residual variance, or it could be due to biased sampling. Shallow earthquakes are distributed globally, while deep earthquakes are restricted to collision zones. Collision zones may be more heterogeneous than other provinces of the Earth.

Rather than giving a pessimistic impression of these results, we should like to remind the reader that the simple model presented here gives an excellent fit to the data, as do the *P*-wave results of GDC. Furthermore, the statistical models derived from both *S* and *P* data are similar in pattern. We should also like to remind the reader that the problems of uneven, structure-dependent sampling, poor resolution of shallow heterogeneity and limited spectral resolution are also (potentially greater) problems in deterministic surface and body wave inversions. The accuracy of the scattering power results are probably of order 20–30 per cent (as estimated from differences due to different damping parameters, different intercept estimates, negative power etc.); but this is to be contrasted with higher precision (from error bars estimated from estimates of error in data), and an order of magnitude difference in power between the upper and lower mantle.

All seismic waves are affected by anisotropy. Our summary ray formulation is largely oblivious to azimuthal anisotropy, since most component rays of summary rays travel the same direction. At the very largest scales though we consider many orientation, hence, a part of the scatter at these scales could be due to azimuthal anisotropy.

6 DISCUSSION

We have inverted the variance of *S*-wave traveltime residuals of bundles of rays to obtain a description of the spectrum of lateral heterogeneity as a function of depth through the mantle. The technique yields robust estimates of the traveltime scattering power (the product of a characteristic scalelength of heterogeneity and the square of the RMS slowness variations). The method also yields estimates of the characteristic scalelength from the spectrum, although with less confidence, and by division, the RMS slowness. By extrapolating the scatter to zero area, we can also estimate the scale-incoherent signal, which is a plausible lower bound on noise in the data.

6.1 Intercepts

The teleseismic *S*-wave intercept estimates are relatively constant as a function of epicentral distance. Similarly the *P*-wave intercepts of GDC are constant over teleseismic distances, but increase dramatically where complexities occur on the traveltime curve of *P*-waves.

The data from deeper source depths have lower intercepts. We suggest that this is partly due to a smaller degree of small-scale structure at depth. This is evidenced in the data as a sharper drop in variance as we approach the origin for the curves corresponding to shallower source depths. There is some evidence of this in the smallest scale data (see Figs 1 and 2). As pointed out by GDC a more subtle effect of small-scale structure arises from the fact that the component rays of a summary ray originate at different source depths inside a given source depth bin. If there is heterogeneity at the scalelength of the path difference, then this will introduce scale-independent scatter. Also, it is reasonable to expect more scatter from shallow events than deep events, since their arrivals are frequently complex and emergent, while deep events generally have simpler impulsive arrivals.

If we assume that the depth variation of the intercept is primarily due to small-scale structure and finite depth binning, then it is reasonable to take the deep intercept estimates as a measure of the non-structure noise in the data. For teleseismic *P*-waves this is $\approx 0.25s^2$ (GDC) and for *S*-waves it is $\approx 4s^2$. Estimates of the structural variance are around $1s^2$ for *P*-waves (GDC) and $8s^2$ for *S*-waves. Hence, we estimate the signal-to-noise ratio of teleseismic ISC *S*-wave data as $\sqrt{8/4} = 1.4$ as opposed to the $\sqrt{1.0/0.25} = 2$ estimate of (GDC) for *P*-waves. If we use logarithmic intercept estimates (Fig. 4b) the signal-to-noise ratio changes to $\sqrt{10/2} = 2.3$ for the *S*-wave data.

6.2 Discussion of seismic velocities

Two striking features of the shear-wave velocity heterogeneity are its concentration in the upper mantle compared to the lower mantle, and its similarity to the results of GDC for *P*-waves. The results suggest that the large-scale slowness fluctuations in the mid-upper mantle are more than an order of magnitude larger than the large-scale fluctuations in the mid-lower mantle. The half-widths of the *P*- and *S*-wave studies are similar throughout the uppermost mantle, and in the lowermost mantle where the *S*-wave half-width is not robust we suggest that the *P*-wave half-width of around 1200 km might be a reasonable estimate; i.e. about 2–3 times larger than the 400–600 km of the upper half of the mantle.

Tanimoto (1990) derived a long-wavelength, whole-mantle shear velocity model using surface waves. He finds a minimum in power at around a depth of 2000 km, with variations of the order of 0.7 per cent peak to peak, i.e., RMS amplitude variations of 0.2 per cent. In his study, the shear velocity is expanded in spherical harmonics to degree 6 (scalelength of order 5000 km). Our results suggests RMS variations of order 0.4 per cent (assuming a half-width of 1200 km from GDC); these are larger as expected given that our model extends to higher harmonic degree.

The distribution of seismicity over the Earth prohibits a global, upper mantle, shear velocity model based on teleseismic first arrival body waves. However, a number of regional shear velocity models have been derived using high-resolution waveform modelling. For instance, Grand & Helmberger (1984) derived a model for the Canadian shield (SNA) and another model (TNA) appropriate for younger oceanic structures and the more tectonic parts of western North America. These differ by 11 per cent at around 100 km depth, which implies RMS slowness variations of 2.5 per cent. Grand & Helmberger (1984) similarly derived a model for old ocean in the North Atlantic (ATL). This differs most from model SNA at depth of $\approx$150 km. These differences correspond to RMS slowness variations of 1.5 per cent. Differences between oceanic, tectonic and shield models decrease linearly from 1 per cent at 220 km to zero at 400 km. These results were confirmed in cross-sections from oceanic and tectonic to shield regions (Graves & Helmberger 1988; Helmberger, Engen & Grand 1985) and in a tomographic analysis of the mantle below the North American plate (Grand 1988). Our results suggest RMS slowness variations of 1.6 per cent, between 70 and 300 km, and 2.9 per cent from 300 to 600 km depth. The difference between our results and these previous regional studies is probably due to the uneven distribution of power across the spectra in these two radial bins which is probably the result of imperfect spectral and radial resolution in our study. Nearly all the power between 300 and 600 km depth is in the low harmonic degrees ($l < 50$), while most of the power between 60 and 300 km depth is in the high harmonic degrees ($l > 50$).

The above high-resolution shear body wave studies were regional. Global comparisons can be made with the global, upper mantle, shear velocity models derived from surface waves. Our RMS slowness estimates (for $l \leq 50$) of 0.006 s km^{-1} correspond to about 3 per cent. This implies 10 per cent peak-to-peak variations as opposed to the 1–3 per cent peak-to-peak variations found in surface wave studies (Dziewonski & Woodhouse 1987; Nataf, Nakanishi & Anderson 1986; Tanimoto 1986, 1987, 1988, 1990; Woodhouse & Dziewonski 1984). This is expected, since the stochastic models extend to higher harmonics.

Free oscillations have also been used to estimate heterogeneity. The study of spheroidal, fundamental mode, and free oscillations by Masters *et al.* (1982) suggests a large quadrupolar pattern in the transition region. Splitting of free oscillation multiplets has also been used to invert for aspherical structure (e.g., Ritzwoller, Masters & Gilbert 1986). They found perturbations in density in the lower mantle of up to 0.35 per cent. Using their correlation of density variations and shear velocity variations they suggest variations of up to 0.9 per cent in S-wave velocities. This is for the longest wavelengths only, and they have poor depth resolution in the lower mantle. They do have the advantage of averaging global volumes, and hence, provide a useful comparison to higher resolution studies, e.g., body wave studies, where potential problems with biased sampling are more prominent. Giardini, Li & Woodhouse (1988), similarly looked at the splitting of long-period normal modes. They show that the splitting functions are consistent with a number of models of heterogeneity in the mantle (Dziewonski 1984; Woodhouse & Dziewonski 1984) provided that $d \ln (V_s)/d \ln (V_p)$ is 2.0–2.5.

The small-scale power is large in the upper 300 km. If we suggest that a reasonable estimate of half-width for these heterogeneities is on the order of 10–50 km as suggested by array studies (Aki 1973; Capon 1974), then we find that the small-scale power is equivalent to velocity heterogeneity of a few per cent for S-waves. This compares favourably with the heterogeneity found in array studies.

Both our results from S-waves and the results of GDC from P-waves suggest that the level of small-scale structure in the lower mantle is nearly indistinguishable from zero. This may not be a true reflection of the real Earth, but rather of wave propagation, i.e., due to intrinsic averaging and wavefront healing because of low-frequency effects. Nolet (1987, p. 10) gives the approximation for the maximum width of a seismic ray, width $\approx \sqrt{(\lambda L/8)}$, where λ is the wavelength ($\approx$25 km), and L is the ray length ($\approx$3700 km when bottoming at the top of the lower mantle and $\approx$8000 km when bottoming in the deep mantle). We estimate the ray width at the top of the lower mantle to be around 110 km (equivalent to harmonic degree 325). The estimate for the deep mantle is 160 km (harmonic degree 155). Note that the period of short-period S-waves (4 s) is about four times longer than the period of short-period P-waves (1 s) (Duda 1971). Hence, from these estimates of ray width we suggest that some but not all small-scale structure (harmonic degree >50) in the lower mantle could be imaged by this technique. Mantle plumes associated with hotspots are estimated to be about 150 km in diameter (Loper & Stacey 1983; White & McKenzie 1989) and hence are probably too small to substantially affect our data.

6.3 Comparison with GDC

A comparison of our S-wave results to the P-wave results of GDC is interesting. Most striking is the consistency in the pattern of the radial variation of the strength of heterogeneity between the two studies. Both show strong heterogeneity in the upper 600 km, decreasing as we approach the lowermost mantle. Both studies have a decrease in the scattering power just above the 670 km discontinuity and an increase just below it. The region above the 670 km discontinuity is poorly resolved (see Fig. 7), and hence heavily damped. It is only the consistency of this feature in both studies that makes it worthy of mention.

Because of the difference in resolution (spectral and depth) of the P- and S-wave data sets, the inversions are damped differently. Thus, a quantitative comparison of the local value of the ratio $d \ln (V_s)/d \ln (V_p)$ may be inappropriate in some regions of the Earth. In particular, the heavy damping of the shallowest layers in the S-wave study, where the P-wave study has relatively more scattering power, distorts any direct quantitative comparison. Also, the S-wave study has a different allocation of power between long- and short-wavelength components compared to the P-wave study, especially between 60 and 540 km depth. The raw data, with and without the intercepts removed, suggest $d \ln (V_s)/d \ln (V_p) \approx 1.8$, if P and S are correlated. Hence, the global average of this ratio integrated along a ray is around 1.8.

Other studies have also found values of $d \ln V_s/d \ln V_p \approx 2$. These include comparison of lower mantle P and S velocity models (Davies & Clayton 1986), comparison of P and S station statics (Doyle & Hales 1967; Davies & Clayton 1986) and the optimal conversion factor to convert a lower mantle P-wave model to an S-wave model to give the best match for free oscillation splitting functions (Giardini *et al.* 1988). Li, Giardini & Woodhouse (1991) constrain $d \ln V_s/d \ln V_p$ to be greater than 1.82 from free oscillation data at the 95 per cent confidence level.

The fundamental point is that traveltime residual variances of order $10\,s^2$ must be modelled for S-waves and of order $1\,s^2$ for P-waves. Hence, if the shear and compressional velocity fields are correlated, this suggests that the ratio is of order 2. Dziewonski & Woodhouse (1987) claim that the S and P velocities are correlated. We cannot prove that the fields are correlated, since we have thrown away all positional information. If they are not correlated, that would be just as interesting, since it would imply the importance of effects other than temperature such as mineralogy or composition. The value of 2 is high compared to the results of experiments on mantle minerals at atmospheric conditions. One explanation proposed for a ratio as high as 2 in the upper mantle is that the mantle is partially molten. Giardini *et al.* (1988) obtained a value of 2.5–3 for this ratio in the lower mantle and mentioned the possibility that the high value for the ratio could be the result of comparing data at different periods; short-period (1 Hz) body waves and long-period free oscillations. We compare data of the same period range; and still get a high ratio. Hence, we suggest that the ratio is not a strong function of period. Anderson (1987) and Agnon & Bukowinski (1989), have suggested that such a high value of the ratio is the natural consequence of high pressure. Reynard & Price (1990) have also addressed the effect of high pressure and find no evidence for a high ratio at lower mantle pressures.

Our S-wave half-widths are similar to the P-wave half-widths of GDC. They are constant through the upper half of the mantle, while they appear to increase in the lower half of the mantle. Given the poor resolution the exact values are questionable, especially in the lower half of the mantle, where the model power (product of half-width and mean square of slowness perturbations) is low. Also,

remember that the upper mantle contains considerable small-scale structure, which is ignored in the half-width estimates. We get half-widths of 350–600 km in the upper half of the mantle, and poorly constrained values in the lower mantle (400–5000 km) that straddle the 1200 km half-width estimate of GDC. The half-widths for *P*-waves (GDC) vary between 450 and 550 km in the upper mantle. Hence, the results are consistent with identical scalelengths of *P* and *S* velocities throughout the mantle.

6.4 Implications for deterministic seismic inversions

These stochastic results have implications for deterministic inversions. First, for inversions using the ISC *P* and *S*-wave data sets, the incoherent variance is an estimate of intrinsic noise. This statistical technique can also measure the structural noise by estimating the variance that is present at the scalelength of the inversion. This information can be used to decide the degree to which to damp a solution. Furthermore, the variance as a function of scale can be computed at various stages in a data reduction process and the reduction of the variance at different scales observed (Zhou 1988). Noise and signal estimates as a function of scale can be used to select appropriate grid sizes and to estimate the data covariance.

The results of our inversions give estimates of spectra as functions of depth. This provides information with which one can design the best model parametrization for the objective in mind. One can then evaluate problems due to aliasing that arise from structures smaller than the smallest scalelength of an inversion. Also, one can estimate problems due to contamination arising from poor resolution and uneven ray coverage.

Both the *P* models of GDC and our *S* models have much more power in the upper mantle than in the lower mantle. This suggests that unless the upper mantle structure is well resolved, slight leakage into the lower mantle could lead to appreciable contamination in the lower mantle. This effect has been investigated synthetically by Gudmundsson (1989) and Gudmundsson & Clayton (1991). He found that lower mantle models are severely affected by upper mantle structures and that traveltime images of the core–mantle boundary are completely obscured. Using 'waves' with large wavelengths (e.g. surface waves, and long-period free oscillations) can alleviate these problems, since they intrinsically average the structure at larger scales. But these generally have poor vertical resolution, and hence, can still lead to contamination. Note that surface-wave modelling requires a crustal correction (Woodhouse & Dziewonski 1984; Nataf *et al.* 1986).

6.5 Temperature variations?

It is interesting to speculate that the large-scale seismic variations are due to temperature variations. Our stochastic inversion for *S*-wave heterogeneity can then be used to infer the amplitude and length scale of temperature fluctuations in the mantle. If the upper mantle is primarily comprised of olivine, then the estimate of the temperature derivative of shear velocity is $dV_s/dT = -0.3 \times 10^{-3}\,\mathrm{km\,s^{-1}\,K^{-1}}$ from laboratory experiments at room temperature and atmospheric pressure (Sumino & Anderson 1984). Their measurements were obtained below the Debye temperature. Isaak, Anderson & Goto (1989a) suggest that for forsterite it decreases from $-0.49 \times 10^{-3}\,\mathrm{km\,s^{-1}\,K^{-1}}$ at 300 K to $-0.69 \times 10^{-3}\,\mathrm{km\,s^{-1}\,K^{-1}}$ at 1700 K. Theoretical calculations suggest that these temperature derivatives increase with pressure (decrease in absolute magnitude), such that temperature variations have a smaller effect on velocities at depth (Isaak, Cohen & Mehl 1989c). Duffy & Ahrens (1990) claim that the absolute magnitude of temperature derivatives of seismic velocities decrease substantially with depth, as derived by comparing the difference between velocities measured on the Hugoniot and predicted on the adiabat. If we assume that the temperature derivatives are similar to the values at atmospheric pressure at upper mantle pressures, then, 2 per cent RMS variations of shear velocity in the upper mantle suggest temperature variations of order 350 K. Using $dV_p/dT = -0.5 \times 10^{-3}\,\mathrm{km\,s^{-1}\,K^{-1}}$, 0.4 per cent variations of compressional velocity suggest temperature variations of order 70 K. The discrepancy between the two estimates is related to the high $d\ln(V_p)/d\ln(V_s) = 2$ ratio discussion above. The explanation proposed to explain this difference in the upper mantle is the presence of melt, which affects shear velocity much more than compressional velocity. Notice, that estimates of small scale structure (harmonic degree >50) of 5 per cent (assuming scalelengths of 40 km) for *P*-waves would lead to mean temperature variations of order 1000 K. Note that the quoted variations are RMS estimates, which are considerably smaller than peak-to-peak variations. It is clear that we need other sources of seismic heterogeneity in addition to temperature in the crust and uppermost mantle.

Unfortunately, there are no similar estimates of the velocity temperature derivatives of perovskite, the primary lower mantle mineral (Jeanloz 1989). Assuming that they are not much different from magnesiowustite (the other significant lower mantle mineral), for which Isaak, Anderson & Goto (1989b) found $dV_p/dT = -0.6 \times 10^{-3}\,\mathrm{km\,s^{-1}\,K^{-1}}$ and $dV_s/dT = -0.45 \times 10^{-3}\,\mathrm{km\,s^{-1}\,K^{-1}}$ at 1800 K, we find that the estimated lower mantle temperature variations are 14 and 70 K (for *P* and *S* velocities respectively), again assuming low-pressure coefficients can apply at high pressure. In contrast, assuming a decrease in the absolute magnitude of the lower mantle temperature derivatives of a factor of ≈8 from the low-pressure coefficients, as suggested by the experiments of Duffy & Ahrens (1990), would lead to the temperature estimates being a factor of 8 higher, i.e. ≈130 and 550 K. A figure of 15 K is arrived at by Hager & Clayton (1989) from compressional body wave tomographic images for the very large-scalelength temperature variations ($l \leq 3$). Since our figure is derived for the structure all the way out to $l = 50$, their estimate is larger.

The assumption that all the variations are due to temperature is questionable. Tanimoto (1989) has shown that continental mantle has seismic velocities different to oceanic mantle due to compositional differences. Unfortunately, the surface wave density kernels are small in the deep mantle, and hence the method cannot be extended deeper into the mantle. One might argue that shallow heterogeneities have a small scalelength since plate tectonics and subduction are their primary source. It is possible that compositional effects are sufficient to affect the correlation

of shear and compressional velocities at intermediate scalelengths.

6.6 Comparisons to convection models

What seismic variations might one expect from a convecting mantle? Jarvis (1985) and Jarvis & Peltier (1986, 1989) considered this for the case of a steady, incompressible, constant viscosity, 2-D, unit aspect-ratio convection. They found the boundary layers to have red spectra and that the spectra became progressively whiter towards the middle of cells. Also, the variations had the largest magnitude at the boundaries. At high Rayleigh numbers they found that the convection becomes more vigorous and the boundary layers thinner. Hence, the spectra become whiter, and stronger.

Towards the top of the mantle we have an increase in heterogeneity and a slight decrease in scalelength. The increased heterogeneity is expected for boundary layers, but we would expect an increase rather than a decrease in scalelength. Our inability to discern changes in characteristic length scales could be due to poor spectral resolution, poor vertical resolution or that this boundary layer is also a site of compositional variations. GDC present similar evidence for the core–mantle boundary. This evidence though, was questioned by Gudmundsson (1989), in a synthetic study that showed that due to poor coverage a similar feature can be reproduced in the inversion if the data are contaminated with realistic noise.

Glatzmaier (1988) presented one of the more complex simulations of mantle convection to date and evaluated spectra of the thermal variations. He considered compressible 3-D convection, with large Rayleigh numbers. He found large variations in behaviour as he increased the Rayleigh number. At $Ra = 10^5$ he found a network of narrow, cold downwellings and broad regions of upwellings, but at $Ra = 10^7$ he found instead hot, thin plumes in a broad region of cold downwelling. His model produced maximum temperature deviations of about 200 K at the largest Rayleigh number. His computations were truncated at harmonic degree 65. At $Ra = 10^5$ he found that the thermal variance decreased by nearly an order of magnitude from its peak at degrees 3–10 out to degree 50. While at $Ra = 10^7$ he found that by harmonic degree 50 the thermal variance had only decreased by a factor of 2. This is qualitatively similar to the results of the much simpler 2-D convection models of Jarvis & Peltier (1986).

The lack of short-scale power in the lower mantle can be explained by the convective features being similar or smaller than the averaging ray width (100 km), and being insufficiently sampled, or that they do not exist and only large-scale features exist. The first two explanations agree with our expectations of very narrow features in a high Rayleigh number, convective system with temperature-dependent viscosity. The third explanation is the opposite and would imply very weak convection, i.e., a low Rayleigh number, possibly due to increasing viscosity with depth (Sammis *et al.* 1977) or decreasing coefficient of expansion (Anderson 1987; Chopelas & Boehler 1989) with depth.

In Fig. 14 we present the logarithm of the spectra versus the logarithm of the harmonic degree. These curves are not linear and hence do not suggest a power law. If we forced the spectra to be fit by a power law with the exponent a

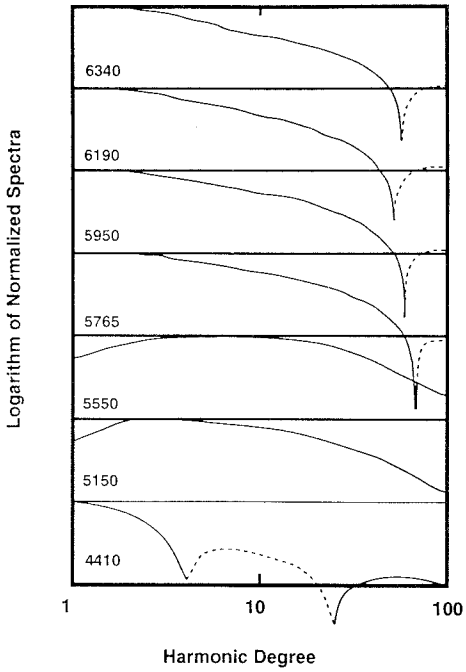

Figure 14. Plot of logarithm of normalized model spectra versus logarithm of harmonic degree. Dashed line represents negative power. The numbers represent the mean radius of each depth bin.

function of scale, it would become more negative as the harmonic degree increased. In contrast, many surface fields are well described by power laws, e.g., heat flow varies like $l^{-0.6}$ (l = harmonic degree). This is much flatter than gravity, which varies like $l^{-2.0}$, or the toroidal or poloidal velocities of the plates, which vary as $l^{-2.3}$, or surface topography, which varies like $l^{-1.2}$ (Kaula 1980).

Glatzmaier (1988) found in his simulations that the logarithm of the thermal variance out to degree 50 is approximately a linear function of the harmonic degree, beyond $l \approx 10$. In Fig. 15 we present the logarithm of the normalized spectrum versus harmonic degree and find a similar linear behaviour. This spectrum has been defined as power per harmonic degree as in Glatzmaier (1988) rather than the conventional power per order [i.e. power per harmonic degree/$(2l + 1)$], and is a depth weighted average through the mantle. Two curves are presented, for the different damping parameters. The equivalent curves for the model resulting from the data with logarithmic intercept estimates removed, and the lifted asymptote data values are nearly identical to the one with a similar damping parameter. The difference between the two curves suggests that we cannot rigorously estimate the slope, but both though do display a nearly linear behaviour between harmonic degrees 10 and 50. If the similarity in form is real, it can potentially provide a constraint on mantle convection and viscosity. A comparison of these curves to Glatzmaier's work would suggest a Rayleigh number of either 10^6 or 10^7.

Comparing the slope of a limited portion of the seismic spectra to the thermal variance of convection calculations is an indirect method to estimate the Rayleigh numb[illegible] and hence, the convective regime of the Earth. It would be reasonable, provided that the thermal variance of convection is a strong function of the Rayleigh number, but

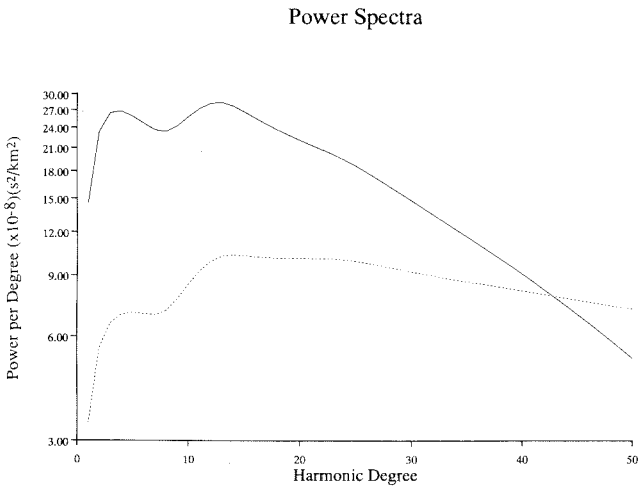

Figure 15. Plot of logarithm of depth averaged spectra of slowness perturbations through the mantle versus harmonic degree on a linear axis. The solid line, (the upper line at low harmonic degree) is for the preferred results (Fig. 9), while the dashed line is for the alternative damping parameters (see Figs 6 and 10). Notice that the spectrum here has been defined as the total power per harmonic degree to allow a direct comparison with Glatzmaier (1988) rather than the more conventional average power per order per harmonic degree. Note that between harmonic degrees 10 and 50 the relationship is linear as is observed in the convection simulations of Glatzmaier (1988).

a weak function of other potential complications (depth and temperature-dependent viscosity, Glatzmaier's simulations are for constant viscosity, plates, etc.). We would also have to assume that the large-scale seismic spectra reflected only temperature variations. If the comparison is reasonable and correct, then the above values would suggest that convection was layered, since a Rayleigh number of order 10^9 would be suggested by whole mantle convection.

7 CONCLUSION

Gudmundsson *et al.* (1990) developed a method to image the spectrum of the Earth's heterogeneity as a function of depth from traveltime data and applied it to ISC *P*-wave data. We have applied the same method to the ISC *S*-wave data. The method yields estimates of incoherent noise in the data, which we find to be about 3 s for shallow events and 2 s for deep events (measured in terms of one standard deviation). The spatially coherent signal in the data is of the order of 3 s. Thus, we estimate a signal-to-noise ratio of slightly larger than unity for the teleseismic, ISC, *S*-wave data set.

We discover that the Earth's seismic heterogeneity is concentrated in the upper 400 km of the mantle. This holds for the *S*-wave results presented in this paper and the *P*-wave results presented by GDC. This could be the result of a decrease in the absolute magnitude of the temperature derivatives of seismic velocity at high pressure or a dramatic deviation away from uniform convection, due to significant radial variations in material properties, e.g., viscosity, coefficient of thermal expansion or thermal diffusivity.

The pattern of the depth variation of the strength of heterogeneity is similar for shear and compressional (GDC) velocities. Assuming that the two velocities are spatially correlated, a signal averaged value of $d \ln (V_s)/d \ln (V_p)$ of at least 2 is required by the data. Note that these data have estimates of noise removed and are from the same period range. This value is appreciably larger than that expected from low-pressure laboratory measurements (0.8–1.4). It is unclear from this study whether this ratio holds for the lower mantle as well as the upper mantle. There is no strong evidence of periodicity in the traveltime autocovariance. Hence, the underlying flow is unlikely to be periodic and the convection is unlikely to be steady. The short length scale at the surface suggests that the upper surface of the Earth is a compositional as well as a thermal boundary layer. The derived spectra have a similar dependence with harmonic degree as the spectra resulting from compressible convection in spherical geometry at Rayleigh numbers of 10^6 and 10^7 (Glatzmaier 1988). As far as the simulation is a reasonable model of mantle convection, and the comparison is legitimate (i.e. seismic heterogeneity is the result of temperature between $l \approx 10$ and 50), then this result favours layered convection.

8 ACKNOWLEDGMENTS

This work was supported by NSF awards EAR83-17623 and EAR83-51371. HD wishes to acknowledge the support of an NERC post-doctoral research fellowship, during the final stages of writing. Malcolm Sambridge and Bruce Buffett read an earlier draft of this paper, which led to significant improvements. Contribution number 5063, Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA 91125, USA.

REFERENCES

Agnon, A. & Bukowinski, M. S. T., 1989. δ_s at high pressure: Is the lower mantle sub-solidus?, *EOS, Trans. Am. geophys. Un.*, **70,** 1212.

Aki, K., 1973. Scattering of P waves under the Montana LASA, *J. geophys. Res.*, **78,** 1334–1346.

Anderson, D. L., 1987. A seismic equation of state 2. Shear properties and thermodynamics of the lower mantle, *Phys. Earth planet. Inter.*, **45,** 307–323.

Capon, J., 1974. Characterization of crust and upper mantle structure under LASA as a random medium, *Bull. seism. Soc. Am.*, **64,** 235–266.

Chopelas, A. & Boehler, R., 1989. Thermal expansion measurements at very high pressure, systematics, and a case for a chemically homogeneous mantle, *Geophys. Res. Lett.*, **16,** 1347–1350.

Clayton, R. W. & Comer, R. P., 1983. A tomographic analysis of mantle heterogeneities from body wave travel times, *EOS, Trans. Am. geophys. Un.*, **62,** 776.

Davies, J. H. & Clayton, R. W., 1986. Lower-mantle S-wave tomography, *EOS, Trans. Am. geophys. Un.*, **67,** 1099.

Davies, J. H., Gudmundsson, O. & Clayton, R. W., 1988. Errors and small scale structure inferred from areal statistics of ISC residuals, *Seism. Res. Lett.*, **59,** 40.

Doyle, H. A. & Hales, A. L., 1967. An analysis of the travel time of S waves to North American stations, *Bull. seism. Soc. Am.*, **57,** 761–771.

Duda, S. J., 1971. Travel times and body wave magnitude, *Pageoph.*, **87,** 13–37.

Duffy, T. S. & Ahrens, T. J., 1990. Sound velocities at high

pressure and temperature and their geophysical implications, preprint.

Dziewonski, A. M., 1984. Mapping the lower mantle: determination of lateral heterogeneity in P velocity up to degree and order 6, *J. geophys. Res.*, **89,** 5929–5952.

Dziewonski, A. M. & Woodhouse, J. H., 1987. Global images of the Earth's interior, *Science,* **236,** 37–48.

Giardini, D., Li, X. & Woodhouse, J., 1988. Splitting functions of long-period normal modes of the Earth, *J. geophys. Res.*, **93,** 13 716–13 742.

Glatzmaier, G. A., 1988. Numerical simulations of mantle convection: time-dependent, three-dimensional, compressible, spherical shell, *Geophys. Astrophys. Fluid Dyn.*, **43,** 223–264.

Grand, S. P., 1988. Tomographic inversion for shear velocity beneath the North American plate, *J. geophys. Res.*, **92,** 14 065–14 090.

Grand, S. P. & Helmberger, D. V., 1984. Upper mantle shear structure of North America, *Geophys. J. R. astr. Soc.*, **76,** 399–438.

Graves, R. & Helmberger, D., 1988. Upper mantle cross section from Tonga to Newfoundland, *J. geophys. Res.*, **93,** 4701–4711.

Gudmundsson, O., 1989. Some problems in global tomography: modeling the core mantle boundary and statistical analysis of travel time data *PhD. thesis,* California Institute of Technology.

Gudmundsson, O. & Clayton, R. W., 1991. A 2-D synthetic study of global traveltime tomography, *Geophys. J. Int.*, **106,** 53–65.

Gudmundsson, O., Davies, H. & Clayton, R. W., 1990. Stochastic analysis of global travel time data: mantle heterogeneity and random errors in the ISC data, *Geophys. J. Int.*, **102,** 25–43.

Hager, B. H. & Clayton, R. W., 1989. Constraints on the structure of mantle convection using seismic observations, flow models, and the geoid, in *Mantle Convection: Plate Tectonics and Global Dynamics,* pp. 657–764, ed. Peltier, W. R., Gordon and Breach Science Publishers, Montreaux.

Helmberger, D. V., Engen, G. & Grand, S., 1985. Upper-mantle cross section from California to Greenland, *J. Geophys.*, **58,** 92–100.

Isaak, D. G., Anderson, O. L. & Goto, T., 1989a. Elasticity of single crystal forsterite measured to 1700 K, *J. geophys. Res.*, **94,** 5895–5907.

Isaak, D. G., Anderson, O. L. & Goto, T., 1989b. Measured elastic constants of single-crystal MgO up to 1800 K, *Phys. Chem. Mineral.*, **16,** 704–713.

Isaak, D. G., Cohen, R. E. & Mehl, M. J., 1989c. Calculated elastic and thermal properties of MgO at high pressures and temperatures, *J. geophys. Res.*, **95,** 7055–7067.

Jarvis, G. T., 1985. The long-wavelength component of mantle convection, *Phys. Earth planet. Inter.*, **40,** 24–42.

Jarvis, G. T. & Peltier, W. R., 1986. Lateral heterogeneity in the convecting mantle, *J. geophys. Res.*, **91,** 435–451.

Jarvis, G. T. & Peltier, W. R., 1989. Convection models and geophysical observations, in *Mantle Convection: Plate Tectonics and Global Dynamics,* pp. 479–594, ed. Peltier, W. R., Gordon and Breach Science Publishers, Montreaux.

Jeanloz, R., 1989. High pressure chemistry of the Earth's mantle and core, in *Mantle Convection: Plate Tectonics and Global Dynamics,* pp. 203–260, ed. Peltier, W. R., Gordon and Breach Science Publishers, Montreaux.

Kaula, W. M., 1989. Material properties for mantle convection consistent with observed surface fields, *J. geophys. Res.*, **85,** 7031–7044.

Li, X.-D., Giardini, D. & Woodhouse, J. H., 1991. The relative amplitudes of mantle heterogeneity in *P* velocity, *S* velocity and density from free-oscillation data, *Geophys. J. Int.*, **105,** 649–657.

Loper, D. E. & Stacey, F. D., 1983. The dynamical and thermal structure of deep mantle plumes, *Phys. Earth planet. Inter.*, **33,** 304–317.

Masters, G., Jordan, T. H., Silver, P. G. & Gilbert, F., 1982. Aspherical earth structure from fundamental spheroidal-mode data, *Nature,* **298,** 609–613.

Nataf, H.-C., Nakanishi, I. & Anderson, D. L., 1986. Measurements of mantle wave velocities and inversion for lateral heterogeneities and anisotropy, III, Inversion, *J. geophys. Res.*, **91,** 7261–7307.

Nolet, G., 1987. Seismic wave propagation and seismic tomography, in *Seismic Tomography with Applications in Global Seismology and Exploration Geophysics,* pp. 1–23, ed. Nolet, G., Reidel, Dodrecht.

Reynard, B. & Price, G. D., 1990. Thermal expansion of mantle minerals at high pressures—a theoretical study, *Geophys. Res. Lett.*, **17,** 689–692.

Ritzwoller, M., Masters, G. & Gilbert, F., 1986. Observations of anomalous splitting and their interpretation in terms of aspherical structure, *J. geophys. Res.*, **91,** 10 203–10 228.

Sammis, C. G., Smith, J. C., Schubert, G. & Yuen, D. A., 1977. Viscosity–depth profile of the Earth's mantle: effects of polymorphic phast transitions, *J. geophys. Res.*, **82,** 3747–3761.

Sumino, Y. & Anderson, O. L., 1984. Elastic constants of minerals, in *Handbook of Physical Properties of Rocks,* pp. 39–137, ed. Carmichael, R. S., CRC Press, Boca Raton, FL.

Tanimoto, T., 1986. The Backus–Gilbert approach to the three dimensional structure in the upper mantle—II. *SH* and *SV* velocity, *Geophys. J. R. astr. Soc.*, **84,** 49–70.

Tanimoto, T., 1987. The three-dimensional shear-wave structure in the mantle by overtone waveform inversion—I. Radial seismogram inversion, *Geophys. J. R. astr. Soc.*, **89,** 713–740.

Tanimoto, T., 1988. The 3-D shear-wave structure in the mantle by overtone waveform inversion—II. Inversion of X-waves, R-waves and G-waves, *Geophys. J. R. astr. Soc.*, **93,** 321–334.

Tanimoto, T., 1991. Waveform inversion for three-dimensional density and S-wave structure, *J. geophys. Res.*, **96,** 8167–8189.

Tanimoto, T., 1990. Long-wavelength *S*-wave velocity structure throughout the mantle, *Geophys. J. Int.*, **100,** 327–336.

White, R. & McKenzie, D., 1989. Magmatism at rift zones: The generation of volcanic continental margins and flood basalts, *J. geophys. Res.*, **94,** 7685–7729.

Woodhouse, J. H. & Dziewonski, A. M., 1984. Mapping the upper mantle: three-dimensional modeling of Earth structure by inversion of seismic waveforms, *J. geophys. Res.*, **89,** 5953–5986.

Zhou, H., 1988. How well can we resolve the deep seismic slab with seismic tomography?, *Geophys. Res. Lett.*, **15,** 1425–1428.